Alljoined1 -A dataset for EEG-to-Image decoding

Jonathan Xu, Bruno Aristimunha, Max Emanuel Feucht, Emma Qian, Charles Liu, Tazik Shahjahan, Martyna Spyra, Steven Zifan Zhang, Nicholas Short, Jioh Kim, et al.

▶ **To cite this version:**

Jonathan Xu, Bruno Aristimunha, Max Emanuel Feucht, Emma Qian, Charles Liu, et al.. Alljoined1 -A dataset for EEG-to-Image decoding. Workshop Data Curation and Augmentation in Medical Imaging at 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, May 2024, Settle, United States. 10.48550/arXiv.2404.05553 . hal-04743819

HAL Id: hal-04743819
https://inria.hal.science/hal-04743819v1

Submitted on 18 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

arXiv:2404.05553v3 [q-bio.NC] 14 May 2024

Alljoined1 - A dataset for EEG-to-Image decoding

Jonathan Xu [*1 2], Bruno Aristimunha [*3 4], Max Emanuel Feucht [*1 5],
Emma Qian [†1], Charles Liu [†1 2], Tazik Shahjahan [†1 2], Martyna Spyra [†1], Steven Zifan Zhang[1,6],
Nicholas Short[1,6], Jioh Kim[1,6], Paula Perdomo[1,6], Ricky Renfeng Mao[1,2], Yashvir Sabharwal[1],
Michael Ahedor[1], Moaz Shoura[6], Adrian Nestor[6]

Abstract

We present Alljoined1, a dataset built specifically for EEG-to-Image decoding. Recognizing that an extensive and unbiased sampling of neural responses to visual stimuli is crucial for image reconstruction efforts, we collected data from 8 participants looking at 10,000 natural images each. We have currently gathered 46,080 epochs of brain responses recorded with a 64-channel EEG headset. The dataset combines response-based stimulus timing, repetition between blocks and sessions, and diverse image classes with the goal of improving signal quality. For transparency, we also provide data quality scores. We publicly release the dataset and all code at https://linktr.ee/alljoined1.

1. Introduction

In the fields of cognitive neuroscience and medical imaging, advancements in deep learning have led to unparalleled precision in decoding brain activity [7, 21, 38–40]. Researchers have translated the intricate patterns of brain activity during various cognitive processes by utilizing neuroimaging modalities, such as functional Magnetic Resonance Imaging (fMRI) and electroencephalography (EEG).

In this context, one particular area of interest is image reconstruction, which involves the decoding of neural responses to visual stimuli, offering insights into how the brain encodes and processes visual information [7, 10, 11, 25, 37, 39].

While fMRI has traditionally been the primary tool for image reconstruction due to its excellent spatial resolution, its low temporal resolution severely delimits actual clinical usage. On the other hand, EEG is a medical modality available in everyday clinical contexts with an excellent time resolution [35, 41, 42]. As neurons fire at millisecond scales, the high temporal resolution provided by EEG is crucial for real-time monitoring of neural dynamics [13, 18, 47]. Additionally, EEG is portable, more accessible to set up, and much more cost-effective than fMRI, making it suitable for real-world applications, including brain-computer interfaces and clinical diagnostics.

The development of very large fMRI-to-image datasets has proven foundational for recent breakthroughs in deep-learning image reconstruction projects. Inspired by the need for such datasets in the EEG domain, we present **Alljoined1**, a novel, large-scale dataset covering a wide range of naturalistic stimuli that allows for robust, generalizable image reconstruction efforts. Our contributions are as follows:

- We propose a stimulus presentation approach that tailors trial duration and session and block repetitions to maximize the signal-to-noise (SNR) ratio.
- We introduce a diverse dataset of EEG responses to 9k unique naturalistic images for each of the eight participants, with 1k additional images shared between participants.
- We perform qualitative comparisons against current EEG-to-image datasets.

2. Related Work

2.1. EEG-to-Image Datasets

EEG-to-image datasets consist of EEG waveforms recorded while participants watch visual stimuli, enabling the study of neural representations in the brain. However, previous research on EEG-based image reconstruction has often relied on datasets exhibiting severe limitations regarding acquisition design or generalizability to naturalistic stimuli [28, 41, 50].

A popular EEG-image dataset is *Brain2Image* [23], which consists of evoked responses to a visual stimulus from distinct image classes. Each block consists of stimuli corresponding only to a single image class. There are 40 classes, with 50 unique images in each class. This dataset has been criticized for having no train-test separation during recording, block-specific stimuli patterns, and lack of

[*] equal contribution. [†] core contribution. [1]Alljoined [2]The University of Waterloo [3]Université Paris-Saclay, Inria TAU, CNRS, LISN [4]Federal University of ABC [5]Vrije Universiteit Amsterdam [6]University of Toronto

consistency across different frequency bands. These factors can incorrectly boost model performance by giving extraneous proxy information about the block rather than the actual image-specific brain responses [4, 29]. An extensive study highlights how many recent EEG-based image reconstruction attempts depend crucially on their block design, demonstrating how similar analytical approaches are not capable of meaningfully decoding EEG signals in a rapid serial visual paradigm (RSVP) [29], even when collecting large amounts of data for only a single subject [3].

Recent studies achieving impressive reconstruction results have relied on data collected with flawed block designs [6, 24, 27], calling the validity of their results into question. As recommended by [4, 29], the stimuli within each block in our dataset were chosen randomly across a variety of natural images, effectively minimizing the risk of block-class correlations.

The diversity of decoding stimuli further limits current EEG-based image reconstruction datasets. While studies like *Brain2Image* or [3] consist of images belonging to 40 classes, several studies utilize a dataset of visual imagery of characters and objects belonging to only 10 different classes, *ThoughtViz* [48]. Such a discretized representation of real world objects fails to account for the continuous, diverse quality of naturalistic stimuli. The same limitation applies to studies utilizing a severely limited quantity of naturalistic stimuli. Approaches to EEG-based image reconstruction derived from the *ThoughtViz* [34, 41], *Brain2Image* [6, 24, 27], or other equally selective datasets [2, 50], may thus suffer from generalizing well to diverse, real-world stimuli. Moreover, a limited number of image classes may encourage image reconstruction models to generate images class-conditionally, rather than reconstructing images based on (continuous) brain-encoded semantic or perceptual attributes of an image.

To account for the diverse and continuous nature of naturalistic images, Alljoined1 consists of 1) 10,000 images per participant 2) that belong to at least one of 80 MS-COCO [31] object categories. Importantly, each MS-COCO category is broader than a single object class (e.g. the *things* category includes car, skateboard, hat, etc.), and each image can belong to up to 5 classes [30].

There are also existing datasets that include naturalistic stimuli, but compromise in other domains. The *MindBigData* initiative [49], or [3] capture a wide sample of images, but are derived only from a single individual, limiting the potential of training image reconstruction models that generalize to other individuals. The *THINGS-EEG1* [17] and *THINGS-EEG2* [14] datasets were acquired using short image presentation times of 50 and 100 ms, and a stimulus onset asynchrony of 100 and 200 ms.

Although the rapid serial visual presentation [16] paradigm proposes disentangling the temporal dynamics of visual processing and categorical abstraction of non-target stimuli, it is not ideal for capturing cortical image processing beyond early visual activity with low noise. We see that [43] obtained the highest accuracy with their EEG-image classifier when focusing on 320-480 ms after stimulus onset, and [36] is able to extract relevant decoding features even around 550 ms after stimulus onset. This suggests that while it takes 50-120 ms for object recognition of a stimulus to register in the visual cortex, a longer stimulus period is beneficial for accuracy on downstream tasks. Alljoined1 consists of extensive data from eight participants, measured with an inter-stimulus interval of 300 ms, which captures important hallmarks in visual processing while maintaining a high presentation frequency [17, 46]. This setup might furthermore allow us to overcome the limitations in decoding image content from EEG activity in RSVP paradigms, as previously reported in [3].

2.2. fMRI-Image Datasets

The recent development of large functional magnetic resonance imaging (fMRI) datasets has enabled researchers to decode and reconstruct images observed by humans with unprecedented accuracy.

The *Brain, Object, Landscape Dataset* (BOLD5000) [9] contains brain responses from 4 human participants who viewed 5,254 images depicting natural scenes from the Scene UNderstanding (SUN) [51], MS-COCO [31], and ImageNet datasets [12]. Similarly, in the *Generic Object Decoding Dataset* (GOD) [22], 1,200 images from the ImageNet database were cropped and shown to 5 participants, resulting in one of the first datasets to establish methods for decoding generic object categories from brain activity.

The *Natural Scenes Dataset* (NSD) [5] consists of the brain responses of 8 human participants passively viewing 9,000–10,000 color natural scenes from MS-COCO. This magnitudes-larger dataset has fueled leaps in reconstruction accuracy seen in recent work like MindEye2 [40]. However, the adaptation of such impressive achievements to real-life contexts is quite limited, as MRI scanners are notoriously expensive and difficult to access.

3. Methods and Materials

3.1. Participants

We collected data from eight participants (six male, two female), with an average age of 22 $\pm$ 0.64 years, all with normal or corrected-to-normal vision, right-handed. All participants were healthy, with no neurocognitive impairments, except 2 participants who reported a history of mental health disorders (e.g. GAD, ADHD). Each participant provided informed consent. The Research Ethics Board approved the procedures as **suppressed for double-blind review**. We note that there are potential limitations of the

study due to the imbalance between the genders of the participants and the low age disparity, which could influence bias and learning with AI models.

3.2. Stimuli

We use the same visual stimuli as what was shown in the fMRI Natural Scenes Dataset (NSD) [5], consisting of $70,566$ images portraying everyday objects and situations in their natural context. All NSD images are drawn from the MS-COCO dataset [30], including annotations about objects and their corresponding category contained in the image. Each image can contain more than one object and more than one object category. These fine-grained object categories are further grouped into *supercategories*, each of which comprehensively includes all related categories as defined subsets.

The current study uses a subset of the first 960 images in the 1000 images shown across all participants in the NSD study. These images are drawn from the *shared1000* subset of the NSD dataset, which comprises 1000 specially curated images that all participants in the original NSD study were presented with [5]. Within this subset of the NSD dataset, the supercategory *person* was most represented, occurring in 50.94% of all images, followed by animal (23.54%) and vehicle (23.33%). The distribution of the supercategories is shown in Figure 1.

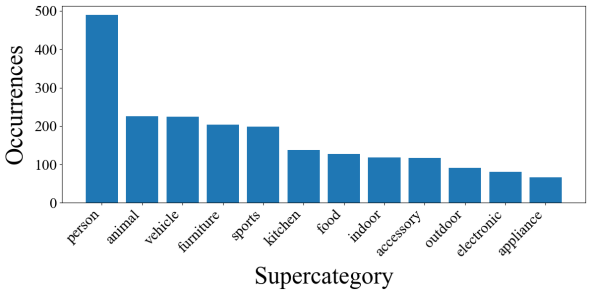


Figure 1. Top 12 most frequently occurring *supercategories* in our dataset.

3.3. Procedure

Images were displayed to participants over the course of multiple one-hour-long sessions. Each session consisted of 16 blocks, wherein images in the first 8 unique blocks were repeated in the second 8 blocks.

The repeated blocks (e.g., blocks 1 & 8, 2 & 9, etc.) contained the same stimuli but in a shuffled order to avoid sequence effects. Within each block, 120 images from the NSD dataset were presented twice, as well as 24 oddball stimuli, amounting to 264 images per block.

Given the within-block and the between-block repetitions of NSD images, each NSD image was presented 4 times to obtain a higher signal-to-noise ratio of the evoked neural responses. Within each trial, an image (NSD or oddball) was presented for 300 ms, followed by 300 ms of a black screen; a white fixation cross was visible on the screen throughout the entire trial.

At the end of each trial, an extra jitter time between 0-50 ms was added for randomness. To ensure focus, participants were prompted to press the space bar when two consecutive trials contained the same image. These oddball trials occurred 24 times within each block; oddballs trials have been discarded from the dataset due to motion artifacts, EEG repetition suppression, and other issues.

3.4. Hardware Setup

We recorded data using a 64-electrode BioSemi ActiveTwo system, digitized at a rate of 512 Hz with 24-bit A/D conversion. The montage was arranged in the International 10-20 System, and the electrode offset was kept below 40 mV. We used a 22 inch Dell monitor at a resolution of 1080p/60Hz to display the visual stimulus. As depicted in Figure 3, the monitor was positioned centrally and placed at a distance of 80 cm to maintain a $3.5°$ visual angle of stimuli. We avoided larger angles to minimize the occurrence of gaze drift.

3.5. Pre-processing

Regarding the dataset pre-processing, we follow recent work on the importance of separating the biomarkers from the central nervous and peripheral systems, as described in [8], and applied the minimum necessary steps. This dataset was pre-processed using the MNE-PYTHON library [15].

Filtering Initially, we applied band-pass filtering with a low frequency of 0.5 Hz and a high frequency of 125 Hz with overlap-add finite impulse response filtering, with range based on [45]. We then apply a notch filter at 60Hz to eliminate power line noise.

Independent Component Analysis (ICA) Next, we performed an ICA decomposition using a FastICA model [1, 19] to separate non-gaussian biological artifacts noise from the signal source. We used a decomposition that retained 95% of the variance and excluded ICs corresponding to eye blinks on the raw data.

Epoching We segmented the continuous data into *epochs*, with each epoch starting at -50 ms onset stimulus and ending at the end of each trial at 600 ms, as described in Section 3.2. The inter-trial jitter periods were excluded from the epochs.

Artifact correction We used the AUTOREJECT algorithm [20] to identify and handle artifact-heavy epochs. Autoreject employs a peak-to-peak threshold criterion separately

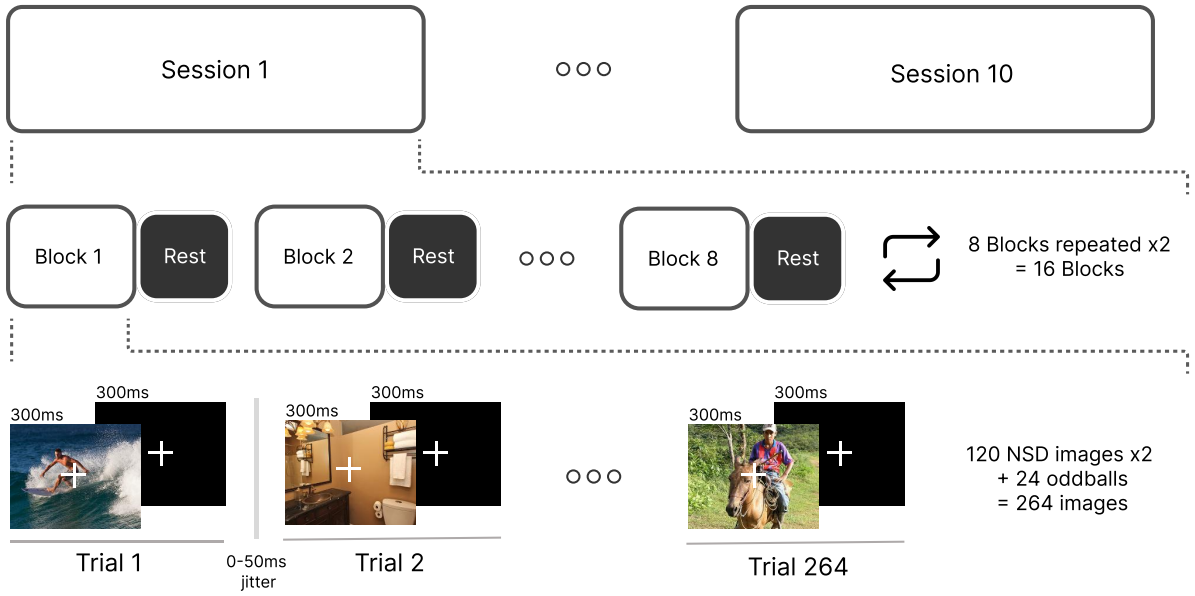


Figure 2. Schematic overview of the structure of trials, blocks, and sessions. Each of the 120 block-specific NSD images is presented twice within each block, and each of the 8 session-specific blocks is presented twice within each session. Each participant performed two sessions on different days. Each of the 10 sessions thus consists of 960 NSD images repeated four times within and across blocks, totaling 9600 unique NSD images per participant.

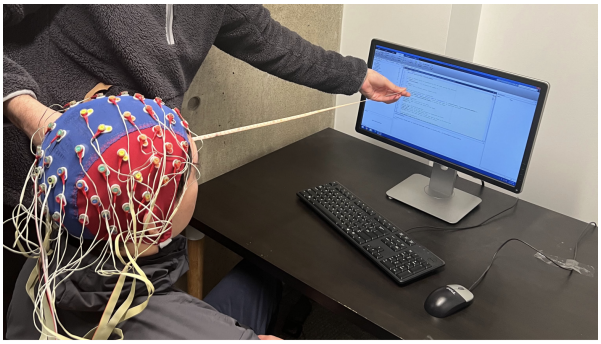


Figure 3. Experimental setup with monitor 80 cm from participant.

for each sensor to determine whether an epoch should be (i) repaired by interpolating the affected sensors using neighboring sensors, or (ii) entirely excluded from further analysis. It performs grid search to determine appropriate values for ρ, the number of channels to interpolate, and κ, the percentage of channels that must agree as a fraction of total channels for consensus. By looking at the number of erroneous sensors per trial, this approach allows correction on a per-trial basis instead of applying a single global threshold to all trials. A mean of 130.75 epochs was dropped per session, with a standard deviation of 260.44.

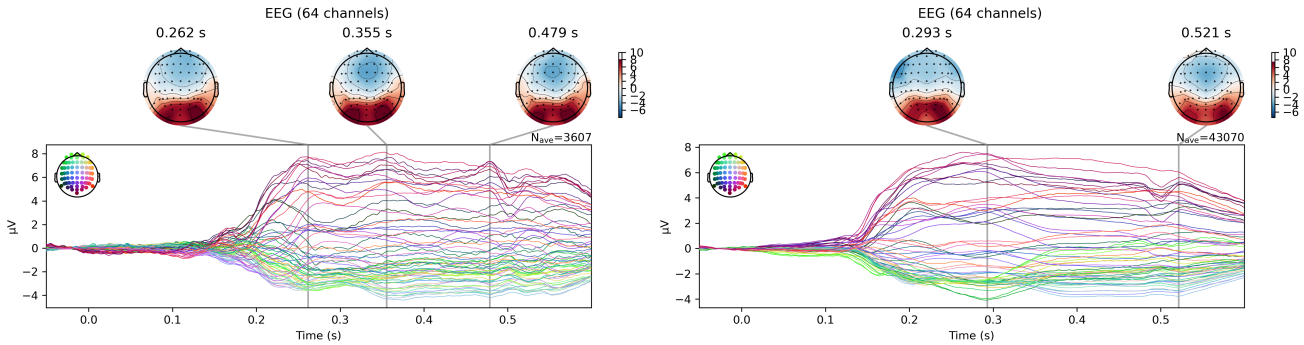
Baseline correction Finally, we re-reference our channels using an average reference scheme, before applying a baseline correction window from -50 ms to 0 ms relative to stimulus onset, following recommendations from [44] for ERP baseline. The epoch data subtracts the average activation during the baseline interval to remove noise from the signal.

4. Analysis

4.1. ERP Analysis

The distribution of event-related potentials (ERPs) across all 64 channels is displayed for a single session of one participant and averaged over all participants and sessions in Figure 5. We observe a strong consistent rise in activity beginning after 150 ms, with a peak between 250 and 300 ms. Note that this latency aligns with the timing parameters of our experimental design, which involves a 300ms presentation followed by a 300ms rest period, with an additional 0-50ms jitter. Both participant- and cohort-level activity exhibits a sustained high level of activity up until 500 ms after stimulus onset, where a consistent dip in activity is observed for both the single participant as well as the whole cohort.

Topographies of activation additionally reveal a strong concentration of positive activation at occipital, parietal, and partially temporal electrode locations and a consistently



(a) First session of the fifth participant

(b) All sessions across all participants

Figure 4. EEG topographic maps and corresponding signals at all 64 electrodes averaged over a) 3823 events for the fifth participant (left) and b) across all sessions for all participants (43070 events) in the Alljoined1 dataset (right), highlighting individual and common brain activity patterns associated with image presentation. An *event* is defined as a specific time point in the experiment.

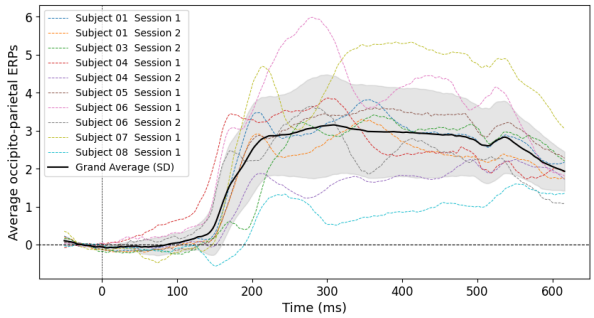


Figure 5. ERPs averaged over occipital and parietal electrodes for all participants and sessions. Shaded areas around the grand average ERP indicate standard deviations at all timepoints.

negative activation at central and frontal areas. This topographical distribution was stable across the duration of the ERP and corresponded well between the single participant and the cohort. Given the strong peak in activity at the occipital and parietal areas, we further investigated the distribution of ERPs across individual participants and sessions at the occipital and parietal electrodes, as displayed in figure 5. While the magnitude of activation differs between participants, we conclude a by-and-large consistent activation pattern across participants and sessions.

4.2. SNR Analysis

The Signal-to-Noise Ratio (SNR) serves as a pivotal metric in evaluating the efficacy of our dataset. To ascertain the SNR, we employ the Standardized Measurement Error (SME) as a gauge for noise assessment [33]. We choose SME as our metric of choice as it is able to robustly quantify the data quality for each participant at each electrode site [32]. The SME is determined by calculating the standard deviation of the aggregated waveform average for each event type across all trials and then dividing this by the square root of the event type's occurrence count. The SNR is subsequently derived by dividing the mean signal values by their corresponding SME.

Figure 6 compares the average SNR across all events in a single session for participant 5 with the average SNR across all events for both sessions concatenated. We see that the SNR is noticeably lower in the multi session graph. This is due to the increased number of repetitions for a given event at different timepoints. This leads to a disproportionately higher standard deviation value and consequently a higher SME and lower SNR. However, that is not to say that the quality of the data is worse. It actually reflects more accurate SNR values as there are more data points, distributed across different sessions. It is also observed that the single sessions graph is more volatile across time, demonstrating a greater variance in SNR values which are captured by having a less accurate metric for noise with less trials to average between. This fluctuation underscores the limited accuracy of noise metrics derived from fewer trials, thus highlighting the critical importance of incorporating repeated measures across sessions or blocks for robust SNR evaluation.

Furthermore, we observe a strong SNR increase 150 ms after stimulus onset. Note that this increase in SNR exhibits the same spiking timing we see in our earlier topographic maps and averaged ERP graphs, suggesting that meaningful activity starts to surface with a considerable delay with respect to stimulus onset.

4.3. Discussion

The ERP and topography analyses, as well as our analysis of the SNR reveal and reinforce several benefits of the acquired dataset, with regard to the stimulus design and timing.

1. **Stimulus duration:** A 300 ms presentation window as well as a subsequent 300 ms rest period allows the capture of both early and late cognitive processes, as ev-

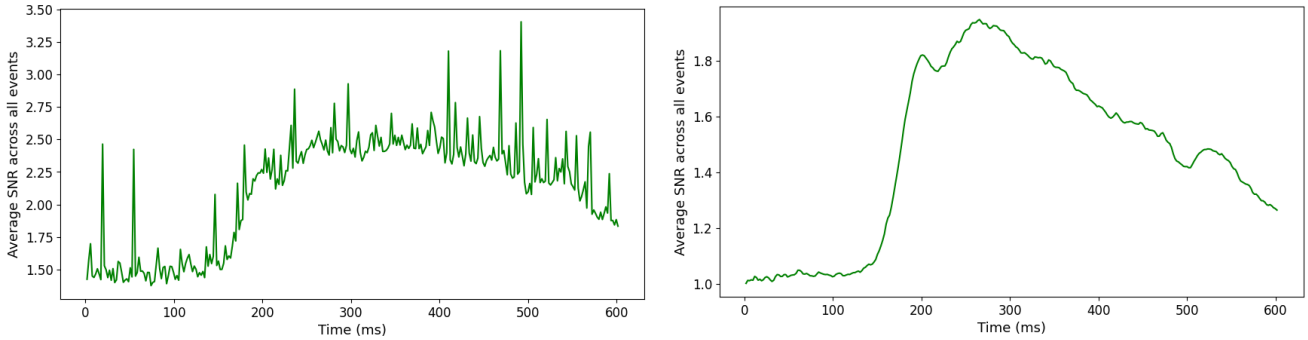


Figure 6. Signal to Noise Rate (SNR) averaged across each session, across each block, and within each block for participant 5. Left: SNR for only the first session 1, Right: SNR for all sessions.

idenced by the single subject peaks at around 262 ms up to 479 ms in Figure 6 a), and the averaged peaks at 293 ms and 521 ms in Figure 6 b), respectively. The duration of 300 ms for image presentation is sufficient for the brain to engage in both perceptual encoding and initial stages of memory processing, which may not be as effectively captured with shorter presentation times. The subsequent 300ms rest period provides a window to measure the brain's higher-level visual and semantic response to the stimuli. The whole ERP thus not only reflects the initial feed-forward transfer of sensory information to visual cortical areas but also the subsequent recurrent interactions involved in attention and semantic analysis, that unfold over hundreds of milliseconds after stimulus onset. The relevance of longer presentation times and longer stimulus-onset asynchrony is additionally supported by the sustained ERP activation presented in Figure 5, as well as the latency of SNR increase and peak in Figure 6.

2. **Comparison with prior studies:** Presentation times of only 100 ms, or stimulus onset asynchronies of only 200 ms fail to capture the rich neural dynamics associated with image processing, involving both lower and higher level processing. In THINGS EEG2 [14], with a shorter 100ms presentation time followed by 100ms of rest, the stimulus exposure may have been insufficient to elicit the full range of cognitive processes to occur. The limited time window could explain the lesser degree of neural activity in the corresponding time window. Similarly, THINGS EEG1 [17] employed a shorter 50ms presentation window followed by 50ms rest, which, while suitable for examining the earliest stages of sensory processing in the visual cortex, likely precluded the phases of the cognitive processes that unfold over a longer period. This includes higher-order mechanisms such as selective attention, working memory updating, and retrieval of semantic associations from long-term memory stores [26].
3. **Phase locking mitigation:** The inclusion of a jitter ranging from 0-50ms helps mitigate phase locking, a phenomenon where the participant's alpha-wave activity becomes synchronously aligned with the pattern of the stimuli after repeated presentations.
4. **Anticipatory bias minimization:** Additionally, the jitter prevents the participants from predicting the exact onset of the next stimulus, thus reducing the potential for anticipatory neural activity that could confound the data.

In conclusion, we choose a 300ms latency as it provides a good trade-off between capturing long-term neural activity whilst maintaining a high presentation frequency. The ideal timing of our experiment ensures the acquisition of a comprehensive ERP waveform, contributing to a more nuanced understanding of cognitive processes and neural dynamics as compared to the shorter intervals used in THINGS EEG2 and THINGS EEG1.

5. Conclusion

We introduce Alljoined1, an EEG-image dataset that uses well-timed stimuli, repetitions between blocks and sessions, and a wide distribution of natural images to create an improved dataset for image decoding tasks. We believe that its size, diversity, and quality will help promote work to better understand the mechanisms of visual processing, and in decoding visual responses in clinical and consumer brain-computer interface (BCI) contexts.

Future directions: We are eager to explore high-density EEG recordings of exclusively the occipital and parietal regions to better target regions of the brain most responsive to visual stimuli. We are also interested in conducting ablation studies on the generalizability of responses to imagined mental imagery. We further believe there is great potential in exploring continuous data collection in natural environments with a wireless headset.

Data availability: Both the raw and preprocessed EEG dataset is available on OSF. Labels to the corresponding NSD image IDs are included in the object files.

Code availability: The stimulus and preprocessing code to reproduce all the results is available on Anonymous GitHub here and here.

6. Acknowledgements

This work was sponsored by Z Fellows, Hack Grants, Moth Fund and Fiona Leng.Bruno's work was supported by DATAIA Convergence Institute as part of the "Programme d'Investissement d'Avenir", (ANR-17-CONV-0003) operated by LISN-CNRS. We would like to thank Dr Sylvain Chevallier for his valuable feedback on this manuscript.

References

[1] Pierre Ablin, Jean-François Cardoso, and Alexandre Gramfort. Faster ICA under orthogonal constraint. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4464–4468. IEEE, 2018. 3

[2] Hajar Ahmadieh, Farnaz Gassemi, and Mohammad Hasan Moradi. Visual image reconstruction based on EEG signals using a generative adversarial and deep fuzzy neural network. *Biomedical Signal Processing and Control*, 87: 105497, 2024. 2

[3] Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Object classification from randomized eeg trials. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3844–3853, 2021. 2

[4] Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Confounds in the Data—Comments on "Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9217–9220, 2022. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2

[5] Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1):116–126, 2022. 2, 3

[6] Yunpeng Bai, Xintao Wang, Yan-pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. Dreamdiffusion: Generating high-quality images from brain eeg signals. *arXiv preprint arXiv:2306.16934*, 2023. 2

[7] Yohann Benchetrit, Hubert Banville, and Jean-Remi King. Brain decoding: toward real-time reconstruction of visual perception. In *The Twelfth International Conference on Learning Representations*, 2024. 1

[8] Philipp Bomatter, Joseph Paillard, Pilar Garces, Jörg Hipp, and Denis Engemann. Machine learning of brain-specific biomarkers from EEG. *bioRxiv*, 2024. 3

[9] Nadine Chang, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific data*, 6(1):49, 2019. 2

[10] Zijiao Chen, Jonathan Xu, Jiaxin Qing, Ruilin Li, and Juan Helen Zhou. Structure-Preserved Image Reconstruction from Brain Recordings. In preparation, 2023. 1

[11] Zijiao Chen, Jiaxin Qing, and Juan Helen Zhou. Cinematic mindscapes: High-quality video reconstruction from brain activity. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[13] Nadine Dijkstra, Pim Mostert, Floris P de Lange, Sander Bosch, and Marcel AJ van Gerven. Differential temporal dynamics during visual imagery and perception. *Elife*, 7: e33904, 2018. 1

[14] Alessandro T. Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M. Cichy. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264: 119754, 2022. 2, 6

[15] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, 7:70133, 2013. 3

[16] Tijl Grootswagers, Amanda K Robinson, and Thomas A Carlson. The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, 188: 668–679, 2019. 2

[17] Tijl Grootswagers, Ivy Zhou, Amanda K Robinson, Martin N Hebart, and Thomas A Carlson. Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, 9(1):3, 2022. 2, 6

[18] Assaf Harel, Iris IA Groen, Dwight J Kravitz, Leon Y Deouell, and Chris I Baker. The temporal dynamics of scene processing: A multifaceted EEG investigation. *Eneuro*, 3(5), 2016. 1

[19] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. Independent component analysis, adaptive and learning systems for signal processing, communications, and control. *John Wiley & Sons, Inc*, 1:11–14, 2001. 3

[20] Mainak Jas, Denis A Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 159: 417–429, 2017. 3

[21] Vinay Jayaram and Alexandre Barachant. MOABB: trustworthy algorithm benchmarking for BCIs. *Journal of neural engineering*, 15(6):066011, 2018. 1

[22] Tomoyasu Horikawa & Yukiyasu Kamitani. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications*, 2017. 2

[23] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. *Brain2Image*: Converting Brain Signals into Images. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1809–1817, Mountain View California USA, 2017. ACM. 1

[24] Nastaran Khaleghi, Tohid Yousefi Rezaii, Soosan Beheshti, Saeed Meshgini, Sobhan Sheykhivand, and Sebelan Danishvar. Visual Saliency and Image Reconstruction from EEG

Signals via an Effective Geometric Deep Network-Based Generative Adversarial Network. *Electronics*, 11(21):3637, 2022. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute. 2

[25] Jean-Rémi King, Laura Gwilliams, Chris Holdgraf, Jona Sassenhagen, Alexandre Barachant, Denis Engemann, Eric Larson, and Alexandre Gramfort. Encoding and Decoding Framework to Uncover the Algorithms of Cognition. In *The Cognitive Neurosciences*. The MIT Press, 2020. 1

[26] Yixuan Ku. Selective attention on representations in working memory: cognitive and neural mechanisms. *PeerJ*, 6:e4585, 2018. 6

[27] Yu-Ting Lan, Kan Ren, Yansen Wang, Wei-Long Zheng, Dongsheng Li, Bao-Liang Lu, and Lili Qiu. Seeing through the Brain: Image Reconstruction of Visual Perception from Human Brain Signals, 2023. arXiv:2308.02510 [cs, eess, q-bio]. 2

[28] Lynn Le, Luca Ambrogioni, Katja Seeliger, Yağmur Güçlütürk, Marcel Van Gerven, and Umut Güçlü. Brain2pix: Fully convolutional naturalistic video reconstruction from brain activity. *BioRxiv*, pages 2021–02, 2021. 1

[29] Ren Li, Jared S. Johansen, Hamad Ahmed, Thomas V. Ilyevsky, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. The perils and pitfalls of block design for eeg classification experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):316–333, 2021. 2

[30] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 2, 3

[31] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 2

[32] Steven J Luck and Emily Kappenman. A new metric for quantifying erp data quality. https://erpinfo.org/blog/2020/4/28/data-quality, 2020. 5

[33] Steven J Luck, Andrew X Stewart, Aaron Matthew Simmons, and Mijke Rhemtulla. Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology*, 58(6):e13793, 2021. 5

[34] Rahul Mishra, Krishan Sharma, R. R. Jha, and Arnav Bhavsar. NeuroGAN: image reconstruction from EEG signals via an attention-based GAN. *Neural Computing and Applications*, 35(12):9181–9192, 2023. 2

[35] Dan Nemrodov, Matthias Niemeier, Ashutosh Patel, and Adrian Nestor. The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction. *Eneuro*, 5(1), 2018. 1

[36] Dan Nemrodov, Shouyu Ling, Ilya Nudnou, Tyler Roberts, Jonathan S. Cant, Andy C. H. Lee, and Adrian Nestor. A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection. *Psychophysiology*, 57(3):e13511, 2020. 2

[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 1

[38] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5): 051001, 2019. 1

[39] Paul Steven Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Cohen Ethan, Aidan James Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth Norman, and Tanishq Mathew Abraham. Reconstructing the Mind's Eye: fMRI-to-Image with Contrastive Learning and Diffusion Priors. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 1

[40] Paul S Scotti, Mihir Tripathy, Cesar Kadir Torrico Villanueva, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A Norman, et al. MindEye2: Shared-Subject Models Enable fMRI-To-Image With 1 Hour of Data. *arXiv preprint arXiv:2403.11207*, 2024. 1, 2

[41] Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, and Shanmuganathan Raman. EEG2IMAGE: Image Reconstruction from EEG Brain Signals, 2023. arXiv:2302.10121 [cs, q-bio]. 1, 2

[42] Prajwal Singh, Dwip Dalal, Gautam Vashishtha, Krishna Miyapuram, and Shanmuganathan Raman. Learning Robust Deep Visual Representations from EEG Brain Recordings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7553–7562, 2024. 1

[43] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6809–6817, 2017. 2

[44] Darren Tanner, James JS Norton, Kara Morgan-Short, and Steven J Luck. On high-pass filter artifacts (they're real) and baseline correction (it'sa good idea) in ERP/ERMF analysis. *Journal of neuroscience methods*, 266:166–170, 2016. 4

[45] Yunzhe Tao, Tao Sun, Aashiq Muhamed, Sahika Genc, Dylan Jackson, Ali Arsanjani, Suri Yaddanapudi, Liang Li, and Prachi Kumar. Gated transformer for decoding human brain EEG signals. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 125–130. IEEE, 2021. 3

[46] Lina Teichmann, Martin N Hebart, and Chris I Baker. Multidimensional object properties are dynamically represented in the human brain. *bioRxiv*, 2023. 2

[47] Simon Thorpe, Denis Fize, and Catherine Marlot. Speed of processing in the human visual system. *nature*, 381(6582): 520–522, 1996. 1

[48] Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. ThoughtViz: Visualizing Human Thoughts Using Generative Adversarial Network. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 950–958, Seoul Republic of Korea, 2018. ACM. 2

[49] David Vivancos and Felix Cuesta. MindBigData 2022 A Large Dataset of Brain Signals. *arXiv preprint arXiv:2212.14746*, 2022. 2

[50] Suguru Wakita, Taiki Orima, and Isamu Motoyoshi. Photorealistic Reconstruction of Visual Texture From EEG Signals. *Frontiers in Computational Neuroscience*, 15, 2021. 1, 2

[51] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356–365, 2016. 2